Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 46 (2015) 449 – 456

Procedia Computer Science

International Conference on Information and Communication Technologies (ICICT 2014)

Identifying knowledge indicators in Higher Education Organization

Preeti Gupta[a,*], Deepti Mehrotra[b], T.K.Sharma[c]

[a,c]*Amity School of Engineering & Technolgy,Amity University Rajasthan,Jaipur,India*
[b]*Amity School of Engineering & Technolgy,Amity University Uttar Pradesh,Noida, India*

Abstract

With the advent of K-economy, all leading business organizations are incorporating knowledge management as an integral part of their functioning. Education sector, though a non-profit domain has also witnessed increase in the implementation of business intelligence and knowledge centric processes over the years. The paper exemplifies the importance of knowledge evaluation in higher education organizations, through measures developed using data mining techniques. Applying these knowledge measures or indicator in the education domain eventually help the higher education organizations to establish themselves as knowledge centric higher education organizations (KCHEO).

Peer-review under responsibility of organizing committee of the International Conference on Information and Communication Technologies (ICICT 2014)

Keywords: Knowledge Management ; Data Mining ; Knowledge Indicators

1. Introduction

Organizations generate vast amount of information on regular basis. To reap long term benefits, they need to continuously evaluate their data base for finding information with high validity and relevance value. Such information can subsequently be graded as knowledge. The generated knowledge is effective in guiding informed decision making in the organization. Measuring the relevance value of knowledge and the initiatives, pertaining to knowledge management in the organization, has thus become imperative to achieve holistic positive impact. [1, 2]

Knowledge indicators or metrics for measuring the relevance of knowledge are to be designed, for defining the extent to which a unit, system, or process possesses a given feature. Comparison of studies between units, groups,

* Preeti Gupta. Tel.: +91-9928369701;
E-mail address: preeti_i@rediffmail.com

Peer-review under responsibility of organizing committee of the International Conference on Information and Communication Technologies (ICICT 2014)
doi:10.1016/j.procs.2015.02.043

time-periods and geographic regions can also be carried out once the knowledge indicators are established. Moreover designing such knowledge indicators will facilitate empirical validation of theories and relationships between concepts. Metrics therefore help us to learn what works and what does not. Thus development of suitable metrics, through which knowledge generated in an organization can be quantified, are the key to progression of research and implementing practice in the area. [3]

Through this paper, method for measuring relevance of knowledge through knowledge indicators in higher education organization is elaborated. Data Mining (DM) techniques are used to establish these indicators.

The paper is organized as follows. The existing techniques for measuring knowledge management (KM) initiatives in an organization are discussed in Section 2. Measuring the KM initiatives in educational sector through the proposed knowledge indicators based on data mining techniques is described in Section 3.Section 4 establishes the importance of the proposed knowledge indicators through implementation of one of the aspect. Section 5 summarizes the importance of these indicators that are based on DM techniques for knowledge centric higher education organizations.

2. Existing techniques for measuring Knowledge Management initiatives

Intellectual Capital (IC) or Knowledge Capital is an increasingly important proposition in an organization. Several methods for measuring Intellectual Capital have been developed after realizing its significance to the organization. Jurczak [4] identified four main groups under which all methods for measuring IC can be divided.

1. DICM-Direct (D) Intellectual (I) Capital (C) Methods (M)
2. MCM-Market (M) Capitalization(C) Methods (M)
3. ROA-Return (R) on (O) Assets (A) Methods
4. SC-Score(S) card (C) Methods

The Scorecard Methods, as compared to financial metrics, create a broader picture of any organization's well being. Different levels of an organization can apply these methods. These methods follow a bottom-up strategy to measure intellectual capital resources of an organization. Government agencies, non profitable organizations, business houses and agencies related to environmental and social causes benefit from these metrics. Some of the scoreboard techniques are listed in Table 1.

Table 1: Models based on Scorecard methods

Measurement Model	Overview
Skandia Navigator [5]	It is a complete model which focuses on human, structural, customer and organizational capital. 91 intellectual based and 73 traditional based measures, determine the Intellectual capital through the analysis of 5 components, financial, customer, processes, renewal and development, human.
Balanced Scorecard [6]	A comprehensive set of performance indicators for strategic management and measurements reflecting organizational missions and strategies.
Intangible asset monitor [7]	People that are considered as organization's profit generators gets primary emphasis.
IC index [2]	Taking past performance into account measures the intellectual capital of the organization.
Value chain scoreboard [8]	The continuous phases of development- Discovery, Implementation and commercialization is taken into account while creating the matrix of non financial indicators.
Human capital intelligence [8]	Sets of human capital indicators are collected and benchmarked against a database.

It has been noticed that because of insufficient measurement systems for adequately measuring qualitative areas, the organizations concentrate more on the financial areas. Hence to exhibit the increase or decrease in the knowledge activity, knowledge indicators are to be designed. According to Robertson [9] these performance indicators determine the status of the project and state whether the project has established a level of satisfaction or not. Four ways are identified to define a performance indicator.

1. A pointer, marking the occurrence of an event.
2. A fraction, denoting how many times an event takes place compared to how many times it could have taken place in the given time period

3. A percentage
4. A Boolean variable, indicating the generation or otherwise, of an event.

3. Measuring Knowledge Management initiatives in education sector

Like other business domains, utilizing knowledge, plays a very crucial role in formulating an effective system in education sector [10-15]. Gupta et.al. [16] have identified, that the important processes that exist in any educational organization are like admission, curriculum development, teaching-learning, conducting examination, research, maintaining alumni relations, strategic planning and many more. Applications and benefits of implementing knowledge centric activities can be very well identified in these processes.

3.1 Proposed work: Designing knowledge indicators for Higher Education Organization

With the advent of privatization in the field of education a growth in the number of private universities and colleges can be witnessed. These universities and colleges like any other organization are also subjected to cutting edge competition. The management is in constant search of excellence to improve the existing processes.

The need is to propose a strategic planning and management tool to align the organization's activities with their vision and strategies. The tool should identify measures and attach some targets so that at a later point in time it is possible to decide whether the organization's performance has met the set expectations or not.

As a part of the work, knowledge indicators have been proposed giving importance to all four pillars of a sound organization. i.e. Human capital, Intellectual capital, Structural capital and Social capital.

In the education system the pillars that have been identified are

- Intellectual accomplishments
- In-house processes
- Stake holders
- Cerebral development and augmentation

The various activities that can be associated with these four pillars of higher education organization are like

- Intellectual accomplishments – like research paper writing, patent filing, study material development, consultancy work undertaken , making students industry ready, molding students for higher education
- In-house processes – like admissions , curriculum development , teaching- learning process
- Stake holders – like Students, Parents, Industry, Society
- Cerebral development and augmentation – initiatives taken for faculty developments, providing aids to facilitate research

Related to the activities associated with these four pillars of the educational organization, measures or knowledge indicators have been proposed.

Further the statistical knowledge indicators are reframed to analyze trends, relationships and associations in the system. Various data mining techniques like rule inductions, classification, regression, are identified to attain the knowledge indicators. Evaluation techniques associated with these data mining techniques are also identified through which the knowledge indicators can be established. The work is compiled in Table 2.

3.2 Establishing knowledge indicators in higher education organization

Over the years, curriculum designing is an important in-house academic process in an educational organization. One of the indicators, for measuring whether the designed curriculum is effective or not, can be established by analyzing the performance of students in the end term examination for a subject. The performance of the students depends on number of factors. If the most effective factor in deciding the performance of the students is identified, curriculum designing can be done in an efficient manner. For the study, the engineering subject of Analog Electronics was chosen. Taking this into consideration the problem statement that was framed was "Identification of the key attribute affecting the result of students in the end term examination for Analog Electronics."

Table 2 Knowledge Indicators for Knowledge Centric Higher education organization (KCHEO)

KCHEO perspective	Measures / Performance Indicators in KCHEO perspectives	Improvement of measure in KCHEO perspectives by introduction of Data Mining technique	DM function to attain the measures	Evaluation methods for DM Tasks
Intellectual accomplishments	Number of research initiatives in terms of research publications, research projects, consultancy works undertaken by the faculty members.	Increase or decline in the overall research culture over the years.	Regression	Slope Analysis
	Percentage of placement of the students on completion of their course work, out of the total no. of students.	Prediction of the placement status of any student during the initial years.	Prediction	Cross Validation
In- House Processes	Passing percentage of the students.	Analyzing the association of various factors affecting student's test score	Supervised Learning through Classification, Prediction Association Rule Mining,	Identification of classifying attribute on the basis of information gain, Confidence, Support and Lift Ratio, Conviction, Succinctness for Association rule Mining
	Attrition rate of students from a course before its completion	Analyzing attrition patterns and predicting drop-out rate in the coming semester	Regression Analysis , Prediction	Cross Validation
Stake Holders (Students/ Parents/ Industry/ Society)	Level of Satisfaction of the students while gaining academic comprehension	Associating the level of satisfaction at various levels with other factors related to the domain.	Association Rule mining, Supervised Learning through Classification,	Confidence, Support and Lift Ratio, Conviction, Succinctness for Association rule Mining
	Level of Satisfaction of the industry in getting market ready and workable individuals			
Cerebral Development and augmentation	No. of opportunities available to faculty members for Enhancement of skill set through regular faculty development programmes.	Identifying non compliance of the requirement	Trend Analysis, Regression	Slope Analysis
	No. of opportunities available to faculty members for promoting research by providing them adequate resources.			

The primary interest of the researcher can be reflected through dependent variable. The identified dependent variable for the study was performance of the student in the end term examination for engineering subject of Analog Electronics. Various independent variables that influences the dependent variable in a positive or negative manner were identified as (1) performance of the student in the prerequisite subject of physics, studied in the previous semester (2) attendance of the student in the subject under scrutiny (3) continuous evaluation grades of the students in the subject under scrutiny (4) practical orientation of the student towards the subject (5) semester grade point average in the previous semester.

Data set used in the study was of second year engineering students and reflected the values pertaining to the various independent variables identified initially. The data set comprised of 41 student records. Missing data was avoided and data was prepared for carrying out further analysis.

Variable(Attribute)	Description	AttributeType	Values	
			Marks obtained	Grade
ContinuousEvaluationMarks	Continuous Evaluation of the student done by the faculty in terms of class test , quiz, assignment submission (Max marks -30)	Nominal	0--5	Fail
			5.1--10	Below Average
			10.1--15	Average
			15.1--20	Good
			20.1--25	Very Good
			25.1--30	Outstanding
EndSemestermarks	End Sem Marks (Max marks -70)	Nominal	0—21	Fail
			21.1--31	Below Average
			31.1--41	Average
			41.1--51	Good
			51.1--61	Very Good
			61.1--70	Outstanding
SGPA_prev	SGPA in previous semester	Nominal	9-10	A
			8-8.9	B
			7-7.9	C
			6-6.9	D
			5-5.9	E
			4-4.9	F
			3-3.9	G
			2-2.9	H
			1-1.9	I
			0-0.9	J
TeachingTeacher	Satisfaction level of the student in terms of teacher teaching the subject	Nominal	A	Good
			B	Fair
Attendance	Attendance	Nominal	75% & Below	Poor
			75.1%-80%	Satisfactory
			80.1%-85%	Fair
			85.1%-90%	Good
			90.1%-95%	VG
			95.1%-100%	Excellent
BaseSubmarks	Marks obtained by the students in the prerequisite subject in the previous semester	Nominal	0—21	Fail
			21.1--31	Below Average
			31.1--41	Average
			41.1--51	Good
			51.1--61	Very Good
			61.1--70	Outstanding
Prac_orient_for_the _sub	Practical orientation of the student towards that subject	Nominal	0—21	Fail
			21.1--31	Below Average
			31.1--41	Average
			41.1--51	Good
			51.1--61	Very Good
			61.1--70	Outstanding

Fig 1: Attributes of the study

As the independent variables influencing the dependent variable in the context of the study were identified from experiential learning it was essential to evaluate and grade the inferred information and identify which of these is the most relevant in the context. i.e which of the identified factors affect the most the performance of the students in the subject of Analog Electronics.

To grade knowledge (effect of one variable on other), the independent variables related to every knowledge deduction were evaluated on the basis of their ability for categorizing the data into various classes related to the end semester marks, in the study. A goodness function was used for this purpose and the variable that produced the "purest" subsets was chosen. The goodness function that is used in the paper is information gain.

Han [17] states, let D be a set consisting of d data samples. In case there are m distinct classes, C_i (for i = 1… m). Let d_i be the number of samples of D in class C_i (for i = 1… m). The expected information needed to classify a given sample is given by $-\sum_{i=1}^{m} p_i \log_2 p_i$ where p_i is the probability that an arbitrary sample belongs to class C_i.
Let attribute A have v distinct values, (a_1, a_2….a_v). Attribute A can be used to partition D into v subsets, (d_1,

$d_2 \ldots d_v$). Let d_{ij} be the number of samples of class C_i in a subset D_j. The entropy, or expected information based on the partitioning into subsets by A is given in equation 1.

$$E(A) = \sum_{j=1}^{v} \frac{d_{1j}+\cdots+dmj}{d} I(d_{1j}, \ldots, d_{mj}) \qquad (1)$$

The smaller the entropy value is, greater the purity of the subset partitions. The information gain, Gain (D, A) of an attribute A, relative to a collection of examples D, is defined in equation 2.

$$\text{Gain}(D, A) = -\sum_{i=1}^{m} p_i \log_2 p_i - \sum_{j=1}^{v} \frac{d_{1j}+\cdots+dmj}{d} I(d_{1j}, \ldots, d_{mj}) \qquad (2)$$

Information gain increases with the average purity of the subsets that a variable produces. Hence the strategy should be to choose variable that result in greatest information gain. On conducting the calculations to know the information gain of the variables the following results were achieved. The results are shown in Table 3.

Table 3: Information Gain of various independent attributes

Variables	**Information Gain**
ContinuousEvaluationMarks	0.496774
BaseSubmarks	0.74
TeachingTeacher	0.04
Prac_orient_for_the_sub	0.41
Attendance	0.69
SGPA_prev	0.665656

It can be witnessed that BaseSubMarks has the highest information gain and hence has the maximum effect on the performance of the student in the subject of Analog Electronics. To further establish the above fact that out of all the independent variables under scrutiny BaseSubmarks is the most important in deciding the final end term grades for the subject of Analog Electronics, the data mining technique of classification was used.

Classification algorithms find a rule or a set of rule to organize data into classes. The traditional and well accepted method of classification is the induction of decision tree. A Decision tree is a rule based – top down induction method which is visually simple and generate production rules. For the study ID3 classifier was used. The popular Open source Data Mining tool of WEKA (Waikato Environment for Knowledge Analysis) was used to develop a decision tree model. Preparation of data included conversion of the data set into .arff format as the native data file format of WEKA is .arff.

The most popular learning method is ID3 family where a decision tree is generated from the cases by recursively selecting the most informative observations whose values separate the cases into relatively homogenous subsets with respect to their solutions.

At a given node the algorithm selects the best attribute according to Quinlan's information gain criterion. The relevance of a selected attribute or test can be explained strategically through the problem solving application of induced decision tree. An execution of the algorithm for a particular scenario provides integrated support for incremental learning and problem solving. The following Decision Tree model (Fig. 2) was obtained using ID3 classifier.

```
BaseSubMarks=Good
| ContinuousEvaluationmarks = Avg : BA
| ContinuousEvaluationmarks = BA : Good
| ContinuousEvaluationmarks = VG : VG
| ContinuousEvaluationmarks = Good : VG
| ContinuousEvaluationmarks = OS : null
| ContinuousEvaluationmarks = F : null
BaseSubMarks=F
| SGPAII=E:F
| SGPAII=G:F
| SGPAII=F
| | ContinuousEvaluationmarks = Avg : F
| | ContinuousEvaluationmarks =  BA:F
| | ContinuousEvaluationmarks = VG : null
| | ContinuousEvaluationmarks = Good : Avg
| | ContinuousEvaluationmarks = OS: null
| | ContinuousEvaluationmarks = F: null
| SGPAII=D
| | ContinuousEvaluationmarks = Avg : BA
| | ContinuousEvaluationmarks =  BA:BA
| | ContinuousEvaluationmarks = VG : null
| | ContinuousEvaluationmarks = Good : VG
| | ContinuousEvaluationmarks = OS: null
| | ContinuousEvaluationmarks = F: null
| SGPAII=B:BA
| SGPAII=C:F
| SGPAII=A
| | ContinuousEvaluationmarks = Avg : F
| | ContinuousEvaluationmarks =  BA:null
| | ContinuousEvaluationmarks = VG : null
| | ContinuousEvaluationmarks = Good : Avg
| | ContinuousEvaluationmarks = OS: null
| | ContinuousEvaluationmarks = F: null
BaseSubMarks=BA
|Attendance=Good:BA
|Attendance=Fair:null
|Attendance=Poor
|  |SGPAII=E:BA
|  |SGPAII=G:null
|  |SGPAII=F:F
|  |SGPAII=D:null
|  |SGPAII=B:null
|  |SGPAII=C:BA
|  |SGPAII=A:null
|Attendance=Poor: null
|Attendance=Excellent
```

Fig 2: Decision Tree through ID3 Algorithm

As BaseSubMarks bears the highest information gain, it can be observed that it appears as the root node of the decision tree. The following knowledge (effect of one variable on the other) can be put to best use for proper curriculum planning, designing adequate lesson plans and evaluation criteria and adoption of suitable pedagogical techniques for improving the overall performance of the subject. Hence information gain can be established as a measure, to grade knowledge acquired by ranking the variables under scrutiny. Through Information gain we were able to compare the various independent variables under scrutiny and their influence on the dependent variable and could use it in a meaningful manner.

4. Conclusion

Due to intangible and vague nature of knowledge resources, the metrics used for measuring the same are quite distinct from others. To adequately access qualitative areas in organization, performance indicators are to be framed. Rather than relying on statistical performance indicators Data Mining techniques can be incorporated to enhance the performance indicators and measure the knowledge activity in the organization. Usage of Data mining techniques for measuring the knowledge activity will help to dig out patterns and establish relationships between variables that are not visible openly.

References

1. Alavi M, Leidner DE. Review: Knowledge Management and Knowledge Management Systems: Conceptual Foundations and Research Issues. *MIS Quarterly* 25(1) 2001; p.107-136
2. KankanHalli A, Tan BCY. A Review of Metrics for Knowledge Management Systems and Knowledge Management Initiatives, in Proceedings of *the 37th Hawaii International Conference on System Sciences* 2004
3. Cook TD, Campbell DT. *Quasi-experimentation: Design and analysis issues for field settings*, Chicago, IL: Rand-McNally; 1979
4. Jurczak J. Intellectual Capital Measurement Methods, *Institute of Organization and Management in Industry ORGMASZ* Vol 1(1); 2008, p. 37 – 45
5. Fairchild AM. Knowledge Management Metrics via a Balanced Scorecard Methodology, in proceedings of the 35th Hawaii International Conference on System Sciences; 2002, p.3173-3180
6. Kaplan R, Norton D. Using the Balanced Scorecard as a strategic management system, *Harvard Business Review*; 2007
7. Sveiby KE. The Intangible Asset Monitor, *Journal of Human Resource Costing & Accounting*, Vol 3 Number 1; 1997, p. 73-97
8. Ramanauskaitė A, Rudžionienė K. Intellectual capital valuation: methods and their classification, *EKONOMIKA* 2013 Vol. 92(2); 2013, p.79-92
9. Robertson J. Metrics for knowledge management and content management, at http://www.steptwo.com.au/papers/kmc_metrics/index. html ; 2003 (Accessed 22 May 2013)
10. Rusli A, Selamat MH, Jaafar A, Abdullah S, Sura S. An Empirical Study of Knowledge Management System Implementation in Public Higher Learning Institution, *International Journal of Computer Science and Network Security;* 2008, p.281-290
11. Sahay A, Mehta K. Assisting Higher Education in Assessing, Predicting, and Managing Issues related to Student Success: A Web-based Software using Data Mining and Quality Function Deployment, in proceedings of *Academic and Business Research Institute Conference*, Las Vegas; 2010, p.1-12
12. Sagsan M. Knowledge Management Discipline: Test for an Undergraduate Program in Turkey. *Electronic Journal of Knowledge Management*; 2009
13. Sharma L, Mehta, N. Data Mining Techniques: A Tool for Knowledge Management System in Agriculture. *International Journal of Scientific and Technology Research*, Vol 1, Issue 5; 2012, p. 67-73
14. Wright H. Tacit Knowledge and Pedagogy at UK Universities; challenges for Effective Management. *Electronic Journal of Knowledge Management*, Vol 6, Issue 1; 2008, p. 49 – 62
15. Cranfield DJ, Taylor J. Knowledge Management and Higher Education: a UK Case Study. *Electronic Journal of Knowledge Management,* Vol 6 Issue 2; 2008, p. 85 – 100
16. Gupta P, Mehrotra D, Singh, R. Achieving Excellence through Knowledge Mapping in Higher Education Institution. *International Journal of Computer Application;* 2012, p. 5-10.
17. Han J, Kamber M. Data Mining: Concepts and Techniques. 2nd ed. Canada: Morgan Kaufmann Publishers; 2000